What is claimed is:

1. A method for factoring an ambiguous finite-state transducer (FST) into an unambiguous FST and a fail-safe FST, comprising the steps of:

preprocessing the ambiguous FST to create a fully-unfolded FST having a plurality of states and arcs, with each arc having at least one input symbol and at least one output symbol;

grouping the plurality of arcs of the fully-unfolded FST into ambiguity fields; and

for each of the plurality of arcs:

if the arc is outside of any ambiguity field, copying the arc to the unambiguous FST, and copying the arc to the fail-safe FST while replacing the corresponding input symbol with the corresponding output symbol; and

if the arc is inside an ambiguity field, copying the arc to the unambiguous FST while replacing the corresponding output symbol with a diacritic, and copying the arc to the fail-safe FST while replacing the corresponding input symbol with the diacritic.

2. The method of claim 1, further comprising the step of factoring the unambiguous FST into a left-sequential FST and a right-sequential FST.

3. The method of claim 1, wherein said preprocessing further comprises the steps of:

concatenating at least one boundary symbol to the ambiguous FST;

minimizing the ambiguous FST to create a minimal FST with an input side and an output side;

left-unfolding the minimal FST to create a left-unfolded FST; and

right-unfolding the left-unfolded FST to create a fully-unfolded FST.

4. The method of claim 3, wherein said preprocessing further comprises:

determining a left-deterministic input finite state automaton by extracting the input side from the minimal FST and determinizing it from left to right;

assigning each state of the left-deterministic input finite state automaton that corresponds to a set of states of the minimal FST a set of state numbers; and copying every state in the minimal FST to the left-unfolded FST as many times as it occurs in different state sets of the left-deterministic input finite state automaton.

5. The method of claim 1, wherein grouping the plurality of arcs into ambiguity fields further comprises grouping the plurality of arcs into disjoint maximal sets of alternative arcs.

6. The method of claim 5, in which arcs grouped together must have:
identical input symbols;
identical sets of input prefixes; and
identical sets of input suffixes.

7. The method of claim 1, wherein the unambiguous FST and the fail-safe FST are adapted for performing language processing.

8. The method of claim 7, wherein the language processing comprises one of tokenization, phonological analysis, morphological analysis, disambiguation, spelling correction, and shallow parsing.

9. The method of claim 1, wherein input prefix and input suffix sets of the states of the fully-unfolded FST are one of identical and disjoint.

10. The method of claim 1, wherein the unambiguous FST and the fail-safe FST are lexical transducers.

11. An apparatus for factoring an ambiguous finite-state transducer (FST) into an unambiguous FST and a fail-safe FST, comprising:

means for preprocessing the ambiguous FST to create a fully-unfolded FST having a plurality of states and arcs, with each arc having at least one input symbol and at least one output symbol;

means for grouping the plurality of arcs of the fully-unfolded FST into ambiguity fields; and

for each of the plurality of arcs:

if the arc is outside of any ambiguity field, copying the arc to the unambiguous FST, and copying the arc to the fail-safe FST while replacing the corresponding input symbol with the corresponding output symbol; and

if the arc is inside an ambiguity field, copying the arc to the unambiguous FST while replacing the corresponding output symbol with a diacritic, and copying the arc to the fail-safe FST while replacing the corresponding input symbol with the diacritic.

12. The apparatus of claim 11, wherein the unambiguous FST and the fail-safe FST are adapted for performing language processing.

13. The apparatus of claim 12, wherein the language processing comprises one of tokenization, phonological analysis, morphological analysis, disambiguation, spelling correction, and shallow parsing.

14. The apparatus of claim 11, wherein the unambiguous FST and the fail-safe FST are lexical transducers.

15. The apparatus of claim 11, wherein input prefix and input suffix sets of the states of the fully-unfolded FST are one of identical and disjoint.